DOCUMENT RESUME

ED 443 852 TM 031 497

AUTHOR Chance, Beth L.
TITLE Components of Statistical Thinking and Implications for Instruction and Assessment.
PUB DATE 2000-04-24
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Curriculum; Definitions; Higher Education; *Instruction; *Statistics; *Student Evaluation; *Thinking Skills

ABSTRACT

This paper focuses on statistical thinking as the third arm of statistical development. The paper opens with a survey of recent definitions of statistical thinking and then attempts to differentiate statistical thinking from statistical literacy and statistical reasoning. Implications for instruction are traced, emphasizing beginning courses for nonstatistics majors. Several suggestions are given to develop habits of statistical thinking in students. The final section of the paper suggests methods and concrete examples for assessing students' ability to think statistically. Literacy can be viewed as understanding and interpreting statistical information presented, and reasoning can be narrowly viewed as working through the tools and concepts in the statistics course. Statistical thinking enables the thinker to move beyond what is taught to question and investigate the issues and data involved. These suggestions are given for students developing the statistical thinking habit: (1) start from the beginning; (2) understand the statistical process as a whole; (3) always be skeptical; (4) think about the variables involved; (5) always relate the data to the context; (6) understand and believe the relevance of statistics; and (7) think beyond the textbook. To develop whether students are applying statistical thinking, problems should be designed to test student reflexes, thought patterns, and creativity in novel situations. (Contains 25 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Components of Statistical Thinking and Implications for Instruction and Assessment

ED 443 852

Beth L. Chance

B. Chance

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM031497

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans
April 2000

BEST COPY AVAILABLE

Presented at Annual Meeting of the American Educational Research Association
New Orleans, April 24, 2000

Components of Statistical Thinking and Implications for Instruction and Assessment

Beth L. Chance (bchance@calpoly.edu)
Statistics Department, Cal Poly, San Luis Obispo, CA 93407

1. Introduction

This paper will focus on the third arm of statistical development: statistical thinking. While having our students "think statistically" is clearly an admirable goal, it's not immediately obvious what this involves and whether or not statistical thinking can be actively taught to our students. Furthermore, what, if any, components of statistical thinking can we expect our beginning students to develop? Thus, this paper will examine the following questions:

- What is statistical thinking?
- How can we teach statistical thinking?
- How do we determine whether students are thinking statistically?

First, the paper provides a survey of recent definitions of "statistical thinking", focuses on elements involved in this process, and attempts to differentiate statistical thinking from statistical literacy and statistical reasoning. Secondly, implications for instruction are given which focus primarily on the beginning courses for non-statistics majors. Several suggestions provide mechanisms for trying to develop "habits" of statistical thinking in students. The final section suggests methods and concrete examples for assessing students' ability to think statistically. While statistical thinking may be distinctly defined, teaching and evaluating thinking greatly overlaps with reasoning and literacy.

2. Definitions of Statistical Thinking

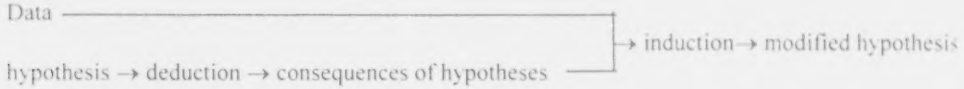
Numerous texts and papers utilize the phrase "statistical thinking" in their title. However, few give a formal definition of statistical thinking. Many appear to use "thinking", "reasoning", and "literacy" interchangeably in an effort to distinguish the thinking and reasoning about statistical concepts from the numerical manipulation that too often characterizes statistical use and instruction. Clearly, we want students to *understand* what they are doing. Current advancements in computing necessitate that "number crunching" no longer dominates the landscape of the introductory course. Instead, we have the luxury of allowing our students to focus on the statistical process that precedes the calculations and the interpretation of the consequences of these calculations.

> Statistical research, practice, and education are entering a new era, one that focuses on the development and use of statistical thinking. (Snee, 1999)

We want students to see the "big picture." However, it has not been as clear how to develop this ability in our students, or even exactly what we mean that big picture to be.

Realizing the inadequacies of current formulations, numerous statisticians and committees have made formal attempts to characterize what is meant by statistical thinking:

In their text, Box, Hunter, and Hunter (1978) outline the process of statistical inquiry through the following schematic:



They encourage statisticians to:

- Find out as much as you can about the problem
- Don't forget nonstatistical knowledge
- Define objectives
- Learn from each other – interplay between theory and practice

Much of this schematic is what researchers are still building on today.

David Moore (1990): The core elements include

1. The omnipresence of variation in processes
2. The need for data about processes
3. The design of data production with variation in mind
4. The quantification of variation
5. The explanation of variation

These ideas were used to form the ASA/MAA joint committee definition (1992):

- the need for data
- the importance of data production
- the omnipresence of variability
- the measuring and modeling of variability

The ASA Working Committing on Statistical Thinking (1993):

a) the appreciation of uncertainty and data variability and their impact on decision making

b) the use of the scientific method in approaching issues and problems

In the domain of quality control and process improvement, Snee (1990) defined statistical thinking as:

> thought processes, which recognize that variation is all around us and present in everything we do, all work is a series of interconnected processes, and identifying, characterizing, quantifying, controlling, and reducing variation provide opportunities for improvement.

The ASQC Glossary of Statistical terms (1996):

> the philosophy of learning and action based on the following fundamental principles:

- all work occurs in a system of interconnected processes
- variation exists in all processes
- understanding and reducing variation are keys to success

In 1998, Mallows argued that the above definitions were missing the "zeroth problem": what data might be relevant. He suggested the following definition:

> the relation of quantitative data to a real-world problem, often in the presence of variability and uncertainty. It attempts to make precise and explicit what the data has to say about the problem of interest.

Mallows also asked whether we can develop a theory of statistical thinking/applied statistics. In 1999, Wild and Pfannkuch attempted just that. Their approach was to ask practicing statisticians and students working on projects what they are "doing" in an attempt to identify the key elements of this previously vague but somehow intuitively understood set of ideas. Their interviews led to development of a four-dimensional framework of statistical thinking in empirical enquiry:

Dimension One: The Investigative Cycle
Dimension Two: Types of Thinking
Dimension Three: The Interrogative Cycle
Dimension Four: Dispositions

They claim that by understanding the thinking patterns and strategies used by statisticians and practitioners to solve real-world problems, and how they are integrated, we will be better able to improve the necessary problem solving and thinking skills in our students. A theme running throughout their article is that the contextual nature of the statistics problem is an essential element and how models are linked to this context is where statistical thinking occurs. While much of the dispositions desired in statistical thinkers, e.g. credulousness and skepticism, is gained through experience, Wild and Pfannkuch further argue that problem solving tools and "worry" or "trigger" questions can be taught to students, instead of relying solely on an apprenticeship model. Clearly, development of the models and prescriptive tools they describe will help with identification of and instruction in statistical thinking.

In a response to Wild and Pfannkuch, Moore argued for "selective introduction" of the types of statistical thinking we introduce to beginning students. In clarifying the "Data, Analysis, Conclusions" portion of the investigative cycle, he argued for the following structure:

> When you first examine a set of data, (1) begin by graphing the data and interpreting what you see; (2) look for overall patterns and for striking deviations from those patterns, and seek explanations in the problem context; (3) based on examination of the data, choose appropriate numerical descriptions of specific aspects; (4) if the overall pattern is sufficiently regular, seek a compact mathematical model for that pattern (p. 251).

For more advanced students he would appear to focus more on issues of measurement and problem formulation as discussed by Mallows. In the same issue, Snee responded that "What data are relevant and how to collect good data are important considerations

and might also be considered core competencies of statisticians" (p. 257) and Smith advocated adding "creativity" as a mode of thinking to Wild and Pfannkuch's list (p. 248).

Following the approach of Wild and Pfannkuch, it seems that a definition of "statistical thinking" includes "what a statistician does." These processes clearly involve, but move beyond, constructing a plot, solving a particular problem, reasoning through a procedure, and explaining the conclusion. Perhaps what is unique to statistical thinking, beyond reasoning and literacy, is the ability to see the process as a whole (with iteration), including "why", to understand the relationship and meaning of variation in this process, to have the ability to explore data in ways beyond what has been prescribed in texts, and to generate new questions beyond those asked by the principle investigator. While literacy can be narrowly viewed as understanding and interpreting statistical information presented, for example in the media, and reasoning can be narrowly viewed as working through the tools and concepts learned in the course, the statistical thinker is able to move beyond what is taught in the course, to spontaneously question and investigate the issues and data involved.

The hope is that by identifying these components, we can attempt to develop them in young statisticians, instead of relying solely on apprenticeship and experience, and also in our beginning students, encouraging them to appreciate this "wider view" (phrase form Wild, 1994) of statistics. In a 1998 newsletter from the University of Melbourne Statistical Consulting Center, Ian Gordon stated: "What professional statisticians have, and amateurs do not have, is precisely that broad view, or overall framework, in which to put a particular problem." Paradoxically, providing a tangible description of this type of insight is very difficult. However, as Wild argues, we may be able to develop "mental habits" that will allow non-statisticians to better appreciate the role and relevance of statistical thinking in future studies. While we may not be able to directly teach students to "think statistically" we can provide them with experiences and examples that foster and reinforce the type of strategies we wish them to employ in novel problems.

3. Implications for Instruction – Developing Habits

These definitions suggest that there is a more global view of the statistical process, including understanding of variability and the statistical process as whole, that we would like to instill in our students. In the past, it was generally assumed that statisticians would develop this manner of thinking through practice, experience, and working with senior statisticians. However, recently there have been more and more calls for novice instruction in the mental habits and problem solving skills needed to think statistically. These mental habits include:

- consideration of how to best obtain meaningful and relevant data to answer the question at hand
- constant reflection on the variables involved and curiosity for other ways of examining and thinking about the data and problem at hand
- seeing the complete process with constant revision of each component
- omnipresent skepticism about the data obtained

- constant relation of the data to the context of the problem and interpretation of the conclusions in non-statistical terms
- thinking beyond the textbook

The question is whether, and how, these habits can be incorporated into beginning instruction. Does the answer vary depending on whether we are talking about courses for statisticians than for other students? Futhermore, where does this component fit into the framework of statistical development?

With current developments in tools for statistical instruction, e.g. case studies, student projects, new assessment tools, it is viable to instill these habits in students. However, the choice of the term "habits" here is quite deliberate, these skills need to be taught through example and repeated use. Furthermore, they don't apply in every situation, but students can learn to approach problems with these general guidelines in mind. Below I begin to outline some of these guidelines and how students can be encouraged to develop these habits. The subsequent section provides suggestions for assessing whether students possess these habits.

3.1 Start from the beginning

Successful statistical consultants have the ability to ask the necessary questions to extract the appropriate data to address the issue in question.

> To me the greatest contributions of statistics to scientific enquiry have been at the planning stage. (Smith, 1999)

Typically it has been assumed that statisticians gain this ability through experience and osmosis. Only by experiencing situations where approaches have failed can we learn how to ask the relevant questions.

As Wild and Pfannkuch (1999) argue, we can provide more structure in this learning process. For example, students need to be given numerous situations where issues of data collection are examined and are clearly relevant to the conclusions drawn from the data. Perhaps the most obvious approach is to ask students to collect data themselves, e.g. measuring the diameter of a tennis ball (Scheaffer, et. al., 1996). Students quickly see the difficulties associated with such a task: Do we have an appropriate measurement tool? What units are we using? How do different methods of measurements contribute to the variability in the measurements? How does variability among observational units affect our results? How do repeated measurements enable us to better describe the "true" measurement? Students clearly see the messiness of actual data collection so often ignored in textbook problems. Students also have a higher degree of ownership and engagement with such assignments.

One of the key questions is "have we collected the right data?" Students can be given numerous examples where "the right answer to the wrong question", often referred to a Type III Error, has led to drastic consequences. The Challenger accident has been held up as an example of not examining the relevant data. Even more simply, students can be asked to compare the prices of small sodas at different Major League Baseball stadiums.

Such data (e.g. as in Rossman and Chance, 2000) should not ignore that the sizes of "small soda" vary from stadium to stadium, and this variation in definition should not be ignored. Or students can compare the percentage of high school students in a state taking the SAT with the average SAT score. Students see that states with lower percentages taking the SAT tend to have higher average scores. They begin to question whether they are looking at the most relevant information to the question.

In my teaching, one way I emphasize to students that all investigations must begin with examination of data collection issues is by moving these topics to be the first discussed in the course. I believe that this emphasizes to students to start with evaluation of the question asked, consideration of other variables, and careful planning of the data collection.

3.2 Understand the statistical process as a whole

Too often, statistical methods are seen as tools that are applied in limited situations. For example, a problem will say "construct a histogram to examine the behavior of these data" or "perform a t-test to assess whether these means are statistically different." This approach allows students to form a very narrow view of statistical application: pieces are applied in isolation as specified by the problem statement. Or a researcher comes to the consulting statistician, data in hand, querying "what method should I use to get the answer I want?" This is extreme, but too often the role of the statistician at the beginning of the investigation is ignored until it is too late.

Instead, instruction should encourage students to view the statistical process in its entirety. Perhaps the most obvious approach is to assign student projects in which students have the primary responsibility of formulating the data collection plan, actively collecting the data, analyzing the data, and then interpreting the data to a general audience. Students are not told which techniques are appropriate, but must decide for themselves, choosing among all topics discussed in the course. Indeed projects have been used with increasing regularity in statistics course and still stand as the best way of introducing students to the entire process of statistical inquiry.

However, as Pfannkuch and Wild (1999) state, "let them do projects" is clearly insufficient as the sole tool for developing statistical problem solving strategies. While we can provide students with such experiences it is also paramount to provide them with a mechanism for learning from the experience and to transfer this new knowledge to other problems. Thus, my students do numerous data collection activities throughout the course and receive feedback that they may apply to their projects. Similarly, they submit periodic project reports during the process to receive feedback on their decisions at each stage. I also structure written assignments where the feedback provided in the grading is expected to be utilized in subsequent assignments. For example, the first writing assignment may ask them to report the mean, median, standard deviation, and quartiles, and comment on the distribution and the interpretations of these numbers. The next assignment merely asks them to describe the distribution, and they are expected to apply their prior knowledge of what constitutes an adequate summary.

These suggestions also encourage students to see the statistical process as iterative. Comments on one project report can be used to modify the proposed procedure before data collection begins. Other approaches that can be used to complement the project component of the course in helping students focus on the overall process include questions at the end of a problem relating back to the data collection issues and how they impact the conclusions drawn. For example, a required component of my student projects is for them to reflect on the weaknesses of the process and suggest changes or next steps for future project teams. Similarly, students can be asked at the end of an inferential question whether the conclusions appear valid based on the data collection procedures.

3.3 Always be skeptical

Wild and Pfannkuch (1999) identified skepticism as a disposition of statistical thinkers that may be taught through experience and seeing "ways in which certain types of information can be unsoundly based and turn out to be false" (p. 235). Research in cognition has demonstrated that to effectively instruct students in a new "way of thinking" they to be given discrediting experiences. Students can be shown numerous examples where poor data collection techniques have invalidated the results. For example, a survey of developers conducted by Microsoft in 1998 at the beginning of their troubles with the Justice Department indicated strong support for integration of the operating system and a network browser. However, closer examination of the poll in court indicated that the questions were "worded in such a way that even market researchers within Microsoft questioned its fairness" (Brinkley, 1999). A follow-up survey showed 44% were in favor of the Department of Justice in contrast to the 85% reported through the initial poll. Further attacks arose when the lawyers produced an email written by Bill Gates in Feb. 1998 that stated "It would HELP ME EMENSELY [sic] to have a survey showing that 90% of developers believe that putting the browser into the operating system makes sense." Through discussion of these examples, students should learn to question the source of the data, the questions used, and the conclusions drawn.

Similar miswordings occurred in a survey which led researchers to conclude that most Americans did not believe the Holocaust had happened (Urschel, 1994). Or the infamous *Literary Digest* poll, whose poor sampling techniques led to an extremely poor prediction of election results. Students need to be exposed to these examples to develop statistical literacy and "worry questions" (Gal et. al, 1995).

Students need to also be given sufficient questions requiring them to choose the appropriate analysis procedure. For example, Short, Moriarty, and Cooley (1995) present a data set on reading level of cancer pamphlets and reading ability of cancer patients. The medians of the two data sets are identical, however, looking at graphs of the two distributions reveals that 27% of the patients would not be able to understand the simplest pamphlet. The authors note that:

> Beginning with the display may 'spoil the fun' of thinking about the appropriateness of measuring and testing centers. We have found that

constructing the display only after discussing the numerical measures of center highlights the importance of simple displays that can be easily interpreted and that may provide the best analysis for a particular problem.

Similarly, no inferential technique should be taught without also examining its limitations. For example, large samples lead to statistical significance *only* in those cases where all other technical conditions are also met. The *Literary Digest* had a huge sample size but the results were still meaningless. Students can be taught to appreciate these limitations and understand when they will need to consult a statistician to determine appropriate methods not covered in their introductory curriculum.

Thus, we can integrate such exposures into instruction instead of only providing problems with nice, neat integer solutions. Through repeated exposure and expectations of closer examination, students should learn to generate these questions on their own, whether they want to or not. I knew I had succeeded when one student indicated that she could no longer watch television, as she was now constantly bombarding herself with questions about sampling and question design. These approaches should help instill the constant skepticism Pfannkuch and Wild (1999) observed in their interviews with professional statisticians.

3.4 Think about the variables involved

Here three issues are paramount: Are they the right variables? How do I think the variable will behave? Are there other variables of importance?

As Mallows (1998) argues, too often we ignore the problem specification in introductory courses, instead starting from the model, assuming the model is correct, and developing our understanding from that point forward. Similarly, Wild and Pfannkuch (1999) argue that we do not teach enough of the mapping between the context and the models. However, particularly in courses for beginning students, these issues are quite relevant and often more of interest to the student. Students are highly motivated to attempt to "debunk" published studies, highlighting areas they feel were not sufficiently examined. This natural inclination to question studies should be rewarded and further developed.

Asking students to reflect on whether the relevant data have been collected was discussed in Section 3.1. Students can also be instructed to always conjecture how a variable will behave (e.g. shape, range of values), before the data have been collected. For example, students can be asked to sketch a graph of measurements of walking time before the data is gathered in class. By anticipating variable behavior, students will better be able to identify unexpected outcomes and problems with data collection. Students will also be able to determine the most appropriate subsequent steps of the analysis based on the shape and behavior of the data. Students also develop a deeper understanding of variation and how it manifests itself in different settings. Students need to be encouraged to think about the problem and understand the problem sufficiently to begin to anticipate what the data will be like.

A statistical thinker is also able to look beyond the variables suggested by the practitioner and guard against ignoring influential variables or drawing faulty causal conclusions. For example, Rossman (1996) presents an example demonstrating the strong correlation between average life expectancy in a country and number of people per television in the country. Too often, people tend to jump to causal conclusions. Here, students are able to postulate other variables that could explain this relationship, such as wealth of country. Similarly, in the SAT example highlighted in Section 3.1, students should consider geography as an explanation for the low percentage of students taking the SATs in some states. Overall, students need to realize that they may not be able to anticipate all relevant variables, highlighting the importance of brainstorming prior to data collection, discussion with practitioners, and properly designed experiments.

3.5 Always relate the data to the context

Students should realize that no numerical answer is sufficient in their statistics course until this answer is related back to the context, to the original question posed. Students should also be encouraged to relate the data in hand to previous experiences and to other outside contexts. Thus, reporting a mean or a p-value should be deemed insufficient presentation of results. Rather, the meaning is given when these numbers are interpreted in context.

For example, data on the weights of the 1996 U.S. Men's Olympic Rowing team contain an extreme low outlier. Most students will recognize that value as the coxswain and will be able to discuss the role of that observation in the overall data summary. Similarly, data on inter-eruption times of the Old Faithful geyser show two distinct mounds, and students can speculate as to the causes of the two types of eruptions. While not all students will possess the outside knowledge needed in each of these settings, these data can be used in classroom discussions to encourage students to always relate their statistical knowledge to other subjects, e.g. geology, biology, psychology, instead of learning statistics and other subjects in "separate mental compartments" (Wild, 1994). These examples also encourage students in "noticing variation and wondering why" (Mullins in Wild and Pfannkuch, 1999).

Another example that highlights to students the importance of the problem context is the "Unusual episode" (e.g. Dawson, 1995). In this example, students are provided with data on number of people exposed to risk, number of deaths, economic status, age, and gender for 1323 individuals. Based solely on these data tables and yes/no questions of the instructor, students are asked to identify the unusual episode involved. This activity encourages students to think about context, hypothesize explanations, and search for meaning, similar to the sleuthing work done by practicing statisticians.

3.6 Understand (and believe) the relevance of statistics

Extending the previous point, students can be instructed to view statistics in the context of the world around them. Techniques range from having students collect data on themselves and their classmates to having students bring in examples of interest from recent news articles. I often include a graded component in my course where students have to discuss some experience they have with statistics *outside of class* during the term.

For example, students may view a talk in their discipline that utilizes statistics, or may be struck by an interesting statement in the media that they now view differently with their statistical debunking glasses on. Thus, students can be led to appreciate the role of statistics in the world around them.

We can also help students see the crucial role statistics and statistical inference play in interpreting information, e.g. that one encounters in popular media. Not only do "data beat anecdotes" (Moore, 1998), but using statistical techniques allows us to extract meaning from data we could not otherwise. Still, issues of variability heavily influence the information we can learn. One lesson I try to impart to my students is the role of sample size in our inferential conclusions – we are allowed to make stronger statements with larger sample sizes and must be cautious of spurious results with small sample sizes. Students can be lead to discover the effect of sample size on p-value by using technology to calculate the p-value for the same difference in population proportions, but different sample sizes (Rossman, 1996). Thus, we cannot determine if two sample proportions are different until we know the sample sizes involved. Similarly, we cannot compare averages, e.g. GPAs of different majors, without knowing the sample sizes and sample standard deviations involved. Statistical methods are necessary to take sampling variability into account before drawing conclusions, and students need to appreciate their role.

At the same time, statisticians *believe* in what they are doing. Before making any conclusion, the statistical thinker immediately asks for the supporting data. I feel I often succeed too well in helping students question conclusions to the point that they never believe any statistical result. The role of randomness in particular is one where the statistical thinker has faith in the outcome and relies on the randomization mechanism, but the novice thinker is untrusting or continues to desire to list and control all variables they can imagine. Again, much of this belief comes from experience, but students can be shown repeatedly what randomization and random sampling accomplish. For example, an exercise in Moore and McCabe (1998) has students pool results from repeated randomization of rats into treatment groups. Students see the long term regularity and equality of the group means prior to treatment and begin to better understand what randomization does and does not accomplish for them. Students should see this idea throughout the course to better understand the "why" of the techniques they are learning.

Students can also be instructed in making sure all statements are supported by the data. For example, in grading their initial lab assignments my most common feedback is "Why, how do you know this is true?" as I insist they support their claims. Many of the above examples are constant reinforcements to make sure students do not make claims beyond what is supported by the data in hand. Casual uses of statistics in sports provide great fodder for unsubstantiated claims. For example, at the start of a NFL playoff game telecast, it was announced that the Tennessee Titans were 11-1 when they won the coin toss to start the game. The statistical thinker immediately looks for the comparison – what was the team's overall record (13-3)? Is this really unusual behavior? The novice merely accepts the data as presented. Students also need to be cautioned against relying excessively on their prior intuitions or opinions. As an example, students can be asked to

evaluate a baseball team's performance based on the average number and standard deviation of errors per game. Often students will respond with their own opinion about the team, ignoring the data presented. With feedback, they can be coached to specify only "what do the data say". Similarly, we can help students learn to jump to the salient point of a problem, instead of meandering in a forest of irrelevant or anecdotal information.

3.7 Think beyond the textbook

The examples given in Section 3.2 (questions that say "construct a histogram to examine the behavior of these data" or "perform a t-test to assess whether these means are statistically different") also highlight the dependency students develop on knowing which section of the book a question comes from. Students learn to apply procedures when directed, but then after the course are at a loss of where to begin when presented with a novel question.

Students need to be given questions that are more open, with suitable development, and encouraged to examine the question from different directions to build understanding. For example, the Old Faithful data mentioned earlier can fail to reveal the bimodal nature of the data with large bin widths. Students should be encouraged to look at more than one visual display. If the ability to explore is an important goal in the course, than this needs to also be built into the assessment. For example, a question on the 1997 AP Statistics exam asked students to choose among several regression models. A question on the 1998 AP exam asked them to produce a histogram from a scatterplot and to comment on features revealed in one display that were much harder to detect in the other. Students blindly following the TI-83 output often did not see as useful a picture as those selecting their own interval limits or using the nature of the data.

To help students choose among inference procedures discussed, I often give them a group quiz where the procedures are listed and they are asked to identify the appropriate procedure based solely on the statement of the research question, considering the number and type of variables involved. This helps students see that the focus is on translating the question of interest, not just the calculations.

4. Assessing Statistical Thinking

The number one mantra to remember when designing assessment instruments is "assess what you value." If you are serious about requiring students to develop the above habits, than you must incorporate follow-up questions into your assessment instruments, whether final exams or performance assessment components.

For example, Wild (1994) claims he is more interested that students ask questions (e.g. in relation to background knowledge, beyond the subject matter) and so usually gives instructions to his graders to "give credit for anything that sounds halfway sensible." Similarly, in my group project grades, students are rewarded as much for the process as the final product. The experience of participating in the project is my main goal, above the sophistication of the final product. This allows students to analyze data using the techniques discussed in the course rather than the sometimes much more complicated but

purely correct approach. Still, students are required to discuss potential biases and other weakness in their current analysis and generate future questions. This encourages students to reflect on the process, critique their own work, realize the limitations of what they have learned, and see how theory differs from practice – all key components of statistical thinking.

Still, much of our assessment must by necessity rely on more traditional exam based questions. Below are some exam questions that I've given in my service courses (usually adapted from other resources) that attempt to assess students' ability to apply the above mental habits.

> *The underlying principle of all statistical inference is that one uses sample statistics to learn something (i.e. to infer something) about the population parameters. Convince me that you understand this statement by writing a short paragraph describing a situation in which you might use a sample statistic to infer something about a population parameter. Clearly identify the sample, population, statistic, and parameter in your example. Be as specific as possible, and do not use any example which we have discussed in class* (from Rossman, 1996).
>
> This problem requires students to demonstrate their understanding of the overall statistical process, at least from the point of data collection forward. Students are required to extract a general approach from the isolated methods learned in the course. The focus is on the big picture rather than a specific technique. They also have to demonstrate their ability to apply their statistical knowledge to answer a question of interest (an individual assessment to complement the group project).

> *Given data on calories for several Chinese foods, students are asked to produce a histogram (using technology) and then*
> *(b) Do you think it is reasonable to use these data to rank the foods from least to most in terms of calorie content? Explain how else you might look at the data if you were interested counting calories.*
>
> In question (b), I'm hoping students will consider the issue of serving size. This serves as a follow-up question to the small soda costs at baseball games examined in class. This approach should be aided by their graph in which egg rolls and soup, the two appetizers, stand out as low outliers. Thus, students are expected to think beyond the statistical method, utilizing context and behavior of the data in their answer.

As part of its twenty-fifth reunion celebration, the Class of '70 of Central University mails a questionnaire to its members. One of the questions asks the respondent to give his or her total income last year. Of the 820 members of the class of '70, the university alumni office has addresses for 583. Of these, 421 return the questionnaire. The reunion committee computes the mean income given in the responses and announces, "The members of the class of '70 has enjoyed resounded success. The average income of class members is $120,000!". Suggest three different sources of bias or misleading information in this result, being explicit about the direction of bias you expect. (From Freedman, et. al., 1978)

In this problem, students have to apply knowledge from several different parts of the course to critique a statement. This tests students' ability to evaluate published conclusions while focusing on issues of data collection (sampling and nonsampling errors) and resistance. Students are asked to address bias, but are not specifically told to focus on sampling design, questionnaire wording, or resistance.

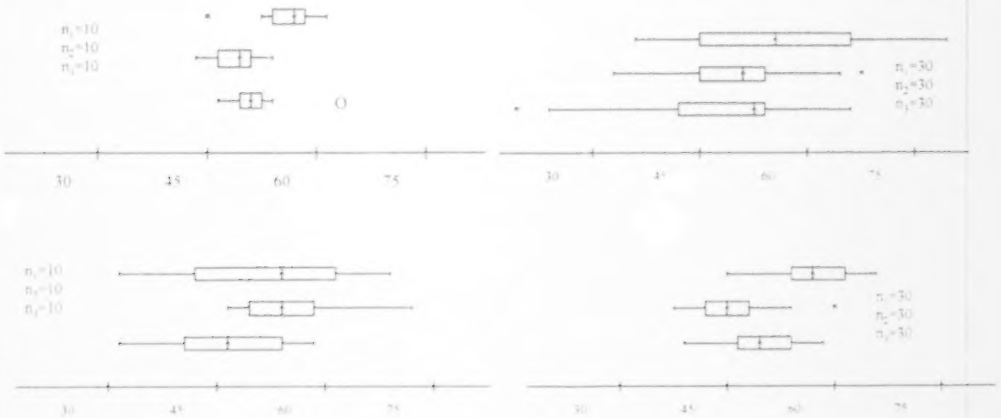
Four (smoothed out) histograms are sketched below. They are histograms for the following variables (in a study of a small town):
(a) Heights of all members of households with children where both parents are less than 24 years old
(b) Heights of both members of all married couples
(c) Heights of all people
(d) Heights of all automobiles.
Match the variables with their histograms. Clearly explain your reasoning from Freedman, et. al., 1978).

This question addresses students' ability to speculate *and justify* different variable behaviors. Students need to think about the context and observational units involved, not just produce graphical displays. Responses are graded on their ability to support their conjecture of the variable behavior.

The FBI reports that nationally 55% of all homicides were the result of gunshot wounds. In a recent sample taken in one community, 66% of all homicides were the result of gunshot wounds. What three possible conclusions can you draw about the percentage from this community compared to the national percentage? What additional information would you need to begin to choose one conclusion over another?

In this short question, the main goal is to see if student understand the role of variability in statistics and why conclusions cannot be drawn until that variation is considered.

A researcher is examining the time for 3 different medicines to register in the blood system (minutes). She wants to test the null hypothesis that the mean times are all the same: H_0: $\mu_1=\mu_2=\mu_3$. For the following four sets of boxplots, order them by smallest p-value to largest p-value and explain your choices. Your grade will be based mostly on your explanation (inspired by Cobb).



Again, this problem does not focus on application of a particular technique but rather asks students to consider issues of sample size and variation in determining statistical significance. Also notice the emphasis on communication for full credit. Thus, students need to understand the purpose and be able to explain the results of the statistical methods. This is similar to the "explain this result to someone who has not taken statistics" question that can be added to the end of a statistical analysis question.

A report based on the Current Population Survey estimates the 1991 median weekly earnings of families of wage and salary works as $664. An approximate 95% confidence interval for the 1991 median weekly earnings of all families of wage and salary workers is $657.14 to $670.86. Interpret this interval. From Moore and McCabe (1998).

This sketch of a problem shows that you can ask students to interpret results from methods not discussed in class. This tests if they can apply the overall reasoning of statistical inference to their interpretation. It addresses the need for students to be able to recognize the relevance of the tools they learn in the course beyond the specific examples (and methods) discussed in class. Furthermore, can students recognize the limitations of the procedures they have learned and when they need to ask for outside consultation?

A university is interested in studying reasons many of their students were failing to graduate. They found that most attrition was occurring during the first three semesters so they recorded various data on the students when they entered the school and their GPA after three semesters. [Students given data set with numerous variables.]
(a) Describe the distribution of GPA for these students.
(b) Is SAT-Math score a statistically significant predictor of GPA for students at this school?
(c) Is there a statistically significant difference between the average GPA values among the majors at this school?
(adapted from Moore & McCabe, 1998)

This type of question is given as a take-home question for the final exam. Students are given one week to identify the relevant statistical methods by reviewing their notes and class examples. Students are instructed to work individually. This type of problem has several goals: can students apply the habits of how to examine a data set numerically and graphically, e.g. shape, center, spread, unusual observations, can students identify and execute the relevant statistical technique with minimal prodding (they don't know what section of the book this question came from so they are missing that context), can they recognize the need for statistical inference to generalize from a sample to a population? With respect to the last point, I have added more and more direction to help students see the need to compute a p-value to attest to "statistical significance." To receive full credit for the inference problems students must still accompany each analysis with appropriate graphical and numerical summaries (again, they must decide what is appropriate). Students are also required to justify their choice of analysis method. To answer these questions, students must decide which variables to examine. This is a complement to giving them a news article and asking them to evaluate the statistical analysis.

While the above questions are aimed primarily at introductory service courses, novice statisticians could be required to analyze the questions like the last in much more depth. For example, with my more mathematically inclined students I expect them to develop a confidence interval formula for a new parameter, e.g. for variances, based on the basic overall structure learned in the course. Chatfield's book, *Problem Solving: A Statistician's Guide*, is an excellent resource for developing further problem solving habits in young statisticians. However, beginning statistics students should also be taught the other mental habits (focus on data collection, question the variables chosen) as well. Our teaching needs to focus "on the big ideas and general strategies" (Moore, 1998). Such instruction will also serve to improve literacy and reasoning:

> students' understanding and retention could be significantly enhanced by teaching the overall process of investigation before the tools, by using tangible case students to introduce and motivate new topics, and by striving for gross (overall) understanding of key concepts (statistical thinking) before fine skills to apply numerical tools." (Hoerl, 1997)

Still, evidence of statistical thinking lies in what students do spontaneously, without prompting or cue from the instructor. Students should be given opportunities to demonstrate their "reflexes." We should see if they demonstrate flexibility in problem solutions and ability to search for meaning with unclear guidelines. These are difficult "skills" to assess and may be beyond what we hope for in the first course for beginning students. However, students can be given more open-ended problems to see how they approach problems on their own and whether they have developed the ability to focus on the critical points of the problem, while still receiving feedback and mentoring from instructors. Recently, "capstone courses" such as this have been incorporated into undergraduate statistics curriculum (e.g. Spurrier) and texts of case studies (e.g. Peck et. al.) have enabled instructors to give students these experiences.

5. Conclusion

Applied to beginning students, I would classify many of the above "habits" as statistical literacy, and this may be all we are hoping to accomplish in many introductory service courses. At this level, I think the types of statistical thinking we aim to teach is what is needed for an informed consumer of statistical information. They serve as the first steps of what we would like to develop in all statisticians, but also what we need to develop in every citizen to understand the importance and need of proper scientific investigation. I imagine that these examples stepped on the toes of statistical reasoning as well, as we encourage students to reason with their statistical tools, and to make sure this reasoning includes awareness of data collection issues and interpretation as well. However, it is through repetition and constant reinforcement that these habits develop into an ingrained system of thought. Through a survey I distributed to students two years after finishing my introductory course, I learned that students often "revert" to some of their old habits. To further develop statistical thinking, these habits need to be continually emphasized in follow-up courses, particularly in other disciplines.

It is also important to remember that when students step into any mathematics course, often they are not expecting to apply their knowledge in these ways. They are accustomed to calculating one definite correct answer that can be boxed and then compared to the numbers in the back of the text. Thus, such habits (questioning, justification, writing in English) require specific instruction and justification in the introductory statistics course. Instructors also need to be aware of the need to allow, even reward, alternative ways of examining data and interpretation.

Thus, we can specifically address the development of statistical thinking in all students. By providing exposure to and instruction in the types of thinking used by statisticians, we can hasten the development of these ways of approaching problems and applying methods in beginning students. These techniques overlap greatly with improving student literacy and reasoning as well. Delving even further into these examples and providing more open-ended problems will continue this development in future statisticians as well. To determine whether students are applying statistical thinking, problems need to be designed that test student reflexes, thought patterns, and creativity in novel situations.

6. References

ASQC (1996). *Glossary of Statistical Terms*, Milwaukee, WI: ASQ.

Box, G.E.P., Hunger, W.G., and Hunger, J.A. (1978). *Statistics for Experimenters*. John Wiley and Sons, New York.

Brinkley, J. (1999). "Microsoft witness attacked for contradictory opinions", *The New York Times*, 15 Jan. 1999, C2.

Dawson, R. J. M. (1995). "The 'Unusual Episode' Data Revisited", *Journal of Statistics Education* [Online], 3(3), http://www.amstat.org/publications/jse/v3n3/datasets.dawson.html

Freedman, D., Pisani, R., Purves, R., & Adhikari, A. (1978). *Statistics*, New York: W.W. North & Company.

Gal, I., Ahlgreen, C., Burrill, G., Landwehr, J., Rich, W. & Begg, A. (1995). "Working group: Assessment of Interpretive Skills", *Writing Group Draft Summaries Conference on Assessment Issues in Statistics Education*, Philadelphia: University of Pennsylvania, 23-35.

Gordon, I. (1998). "From the Director", *News and Views*, (13). http://www.scc.ms.unimelb.edu.au/news/n13.html

Hoerl, R.W. (1997). "Introductory Statistical Education: Radical Redesign is Needed, or is it?" *Newsletter for the Section on Statistical Education of the American Statistical Association*,3(1).

Mallows, C. (1998). "The Zeroth Problem", *The American Statistician*, 52(1), 1-9.

Moore, D.S. (1999). "Discussion: What Shall We Teach Beginners?", *International Statistical Review*, 67(3), 250-252.

Moore, D.S. (1998). "Statistics Among the Liberal Arts", *Journal of the American Statistical Association*, 93(444), 1253-1259.

Moore, D.S. (1990). "Uncertainty" in *On the Shoulders of Giants*, e.g L.A. Steen. National Academy Press, 95-173.

Moore, D.S. and McCabe, G.P. (1998). *Introduction to the Practice of Statistics*. New York: W.H. Freeman and Company.

Peck, R., Haugh, L.D., Goodman, A. eds. (1998). *Statistical Case Studies: A Collaboration Between Academe and Industry*. American Statistical Association/SIAM.

Rossman, A.J. (1996). *Workshop Statistics: Discovery with Data*, New York: Springer-Verlag Publishers.

Rossman, A.J. and Chance, B.L. (2000). *Workshop Statistics: Discovery with Data and Minitab*, Emeryville: Key College Press.

Scheaffer, R., Gnanadesikan, M., Watkins, A., and Witmer, J. (1996). *Activity-Based Statistics*, New York: Springer-Verlag Publishers.

Short, T. H., Moriarty, H., and Cooley, M. E. (1995). "Readability of Educational Materials for Patients with Cancer", *Journal of Statistics Education* [Online], 3(2). http://www.amstat.org/publications/jse/v3n2/datasets.short.html

Smith, T.M.F. (1999). "Discussion" in response to Wild and Pfannkuch, *International Statistical Review*, 67(3), 248-250.

Snee, R.D. (1999). "Discussion: Development and Use of Statistical Thinking: A new Era", *International Statistical Review*, 67(3), 255-258.

Snee, R.D. (1990). "Statistical Thinking and Its Contribution to Total Quality", *The American Statistician*, 44(2), 116-121.

Spurrier, J. D. (1999). *The Practice of Statistics: Putting the Pieces Together*, Belmont, CA: Duxbury Press.

Urschel, J. (1994). "Putting a reality check on 'Holocaust denial'", *USA Today*, January 12, 1994.

Wild, C.J. (1994). "Embracing the 'Wider View' of Statistics", *The American Statistician*, 48(2), 163-171.

Wild, C.J. and Pfannkuch, M. (1999). "Statistical Thinking in Empirical Enquiry", *International Statistical Review*, 67(3), 223-265.